Chapter 6. TRANSCRIPTION, TRANSLATION AND TRANSPORT

6.1 Introduction to transcription, translation and transport
6.2 Transcription of virus genomes
6.3 Transcription in eukaryotes
6.4 Translation in eukaryotes
6.5 Transport in eukaryotic cells
6.6 Transcription and translation in bacteria

Learning outcomes

By the end of this chapter you should be able to

1. explain how virus genes are transcribed and translated;
2. describe the post-translational modifications that some virus proteins undergo;
3. highlight differences in transcription and translation between prokaryotic and eukaryotic cells;
4. discuss the transport of virus proteins and RNA within cells.

6.1 Introduction to transcription, translation and transport

You can probably think of several words beginning with 'trans-' and it is probable that most of them are derived from the Latin trans, meaning across. There are many 'trans' words used in virology and more broadly in biology, including transmission (Chapter 4), transposon (Chapter 20) and transformation (Chapter 22). The definitions of these words involve something going across.

In this chapter we deal with

- transcription = writing across
- translation = bearing across
- transport = carrying across

For our purposes, transcription refers to the writing across of genetic information from a sequence of bases in a nucleic acid to the complementary sequence in messenger RNA (mRNA), while translation converts the genetic information from the language of bases in nucleic acids to the language of amino acids in proteins. Transcription and translation are steps 3 and 4 of our generalized replication cycle.

We also discuss in this chapter the transport of virus proteins and RNAs to particular locations in infected cells. We start with an overview of virus transcription, and then we discuss these three 'trans' processes in eukaryotic cells. At the end of the chapter we point out some aspects of the processes that are different in bacterial cells.

6.2 Transcription of virus genomes

We have seen how there are four main categories of virus genome: dsDNA, ssDNA, dsRNA and ssRNA (Section 3.2). Because of distinct modes of transcription within the dsDNA and ssRNA categories a total of seven classes of viruses can be recognized (Figure 6.1).

This division of the viruses into classes based on genome type and mode of transcription was first suggested by David Baltimore and this scheme of virus classification is named after him. He initially proposed six classes.

In the summary of the scheme depicted in Figure 6.1 most of the nucleic acid strands are labelled (+) or (–). This labelling is relative to the virus mRNA, which is always designated (+). A nucleic acid strand that has the same sequence as mRNA is labeled (+) and a nucleic acid strand that has the sequence complementary to the mRNA is labelled (–).

The viruses with (+) RNA genomes (Classes IV and VI) have the same sequence as the virus mRNA. When these viruses infect cells, however, only the Class IV genomes can function as mRNA. These viruses are commonly referred to as plus-strand (or positive-strand) RNA viruses. The Class V viruses are

commonly referred to as minus-strand (or negative strand) RNA viruses. Class VI viruses must first reverse transcribe their ssRNA genomes to dsDNA before mRNA can be transcribed. Because they carry out transcription in reverse (RNA to DNA) Class VI viruses are known as retroviruses. The ability of some DNA viruses to carry out reverse transcription was discovered later; these viruses became known as pararetroviruses and Class VII was formed to accommodate them.

There are a few single-stranded nucleic acids of viruses where there is a mixture of (+) and (–) polarity within the strand, in other words there are open reading frames (ORFs) in both directions. Genomes of this type are known as ambisense, a word derived from the Latin *ambi*, meaning 'on both sides' (as in ambidextrous). Examples of ambisense genomes include the ssDNA genomes of the geminiviruses, which are plant viruses, and the ssRNA genomes of the arenaviruses, which are animal viruses and include the causative agent of Lassa fever.

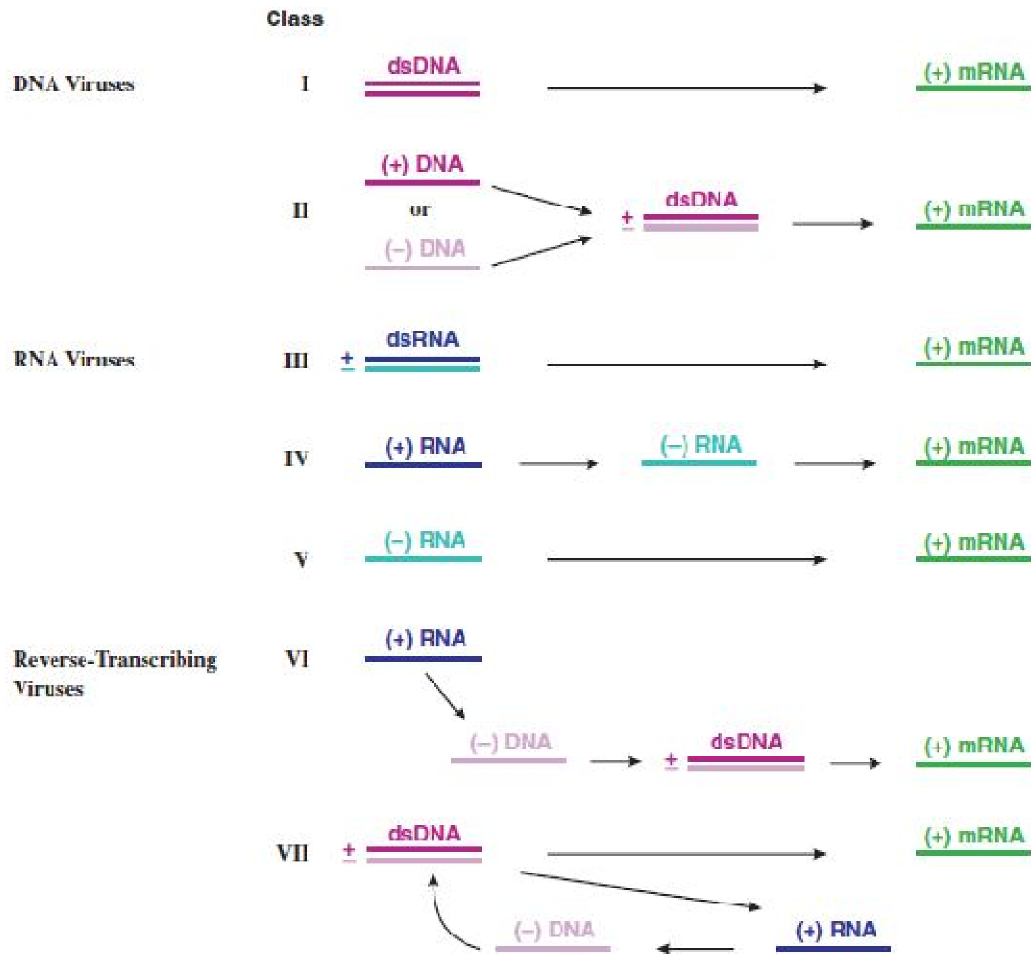

Figure 6.1 Transcription of virus genomes. (+) RNA and (+) DNA have the same sequence as the mRNA (except that in DNA thymine replaces uracil). (−) RNA and (−) DNA have the sequence complementary to the mRNA (except that in DNA thymine replaces uracil). (+) and (−) strands are not indicated for the dsDNA of the Class I viruses as the genomes of most of these viruses have open reading frames (ORFs) in both directions. (+) and (−) strands are indicated for the ssDNA of the Class II viruses. Most of these viruses have either a (+) or a (−) strand genome. A (+) RNA genome (dark blue) has the same sequence as the corresponding mRNA (green). The molecules are shown in different colours to indicate their different functions. In Class VII viruses the (+) RNA shown in blue (pregenome RNA) functions as a template for DNA synthesis (Section 18.8.6). Some of the DNA is used as a template for further transcription. Some ssDNA viruses and some ssRNA viruses have ambisense genomes. This means that that the polarity of the genome is part (+) and part (−).

6.2.1 Modifications to the central dogma

In 1958 Francis Crick proposed a 'central dogma of molecular biology'. JamesWatson, Crick's collaborator in deducing the structure of DNA, made significant contributions to the formulation of the dogma, which stated that the flow of genetic information is always from DNA to RNA and then to protein, with genetic information transmitted from one generation to the next through copying from DNA to DNA (Figure 6.2(a)). Increasing understanding of how viruses replicate their genomes necessitated some modifications to this dogma in 1970; many viruses have RNA genomes that are copied to RNA, and some viruses copy from RNA to DNA (Figure 6.2(b)).

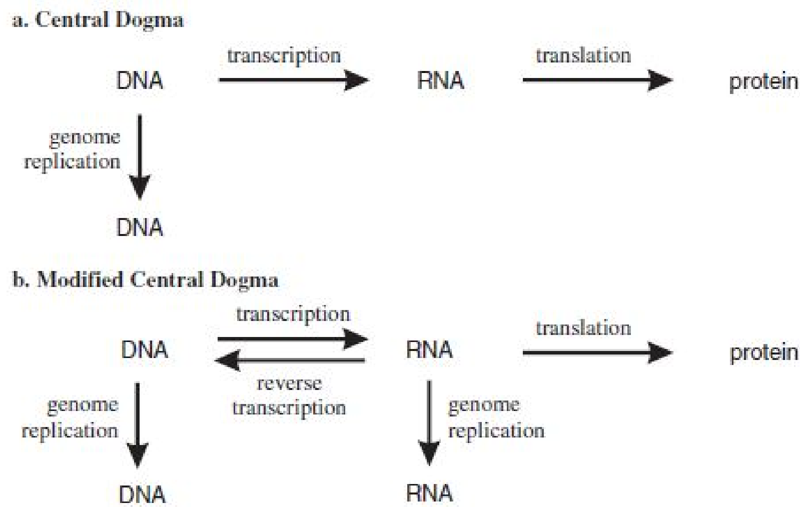

Figure 6.2 (a) Francis Crick's central dogma of molecular biology, which proposed that genetic information is transmitted from DNA to RNA to protein, and from DNA to DNA. (b) Modifications to the central dogma, required because of the various modes of virus transcription and genome replication.

6.3 Transcription in eukaryotes

We start this section with a brief summary of transcription from eukaryotic cell genes, as many viruses transcribe their genes by similar processes, some of them using parts of the cell transcription machinery (Figure 6.3). The expression of a gene is controlled by various sequences in the DNA:

- enhancers – sequences that contain binding sites for transcription factors, which affect the rate of transcription;
- a promoter – the 'on' switch;
- a terminator – the sequence that causes the enzyme to stop transcription.

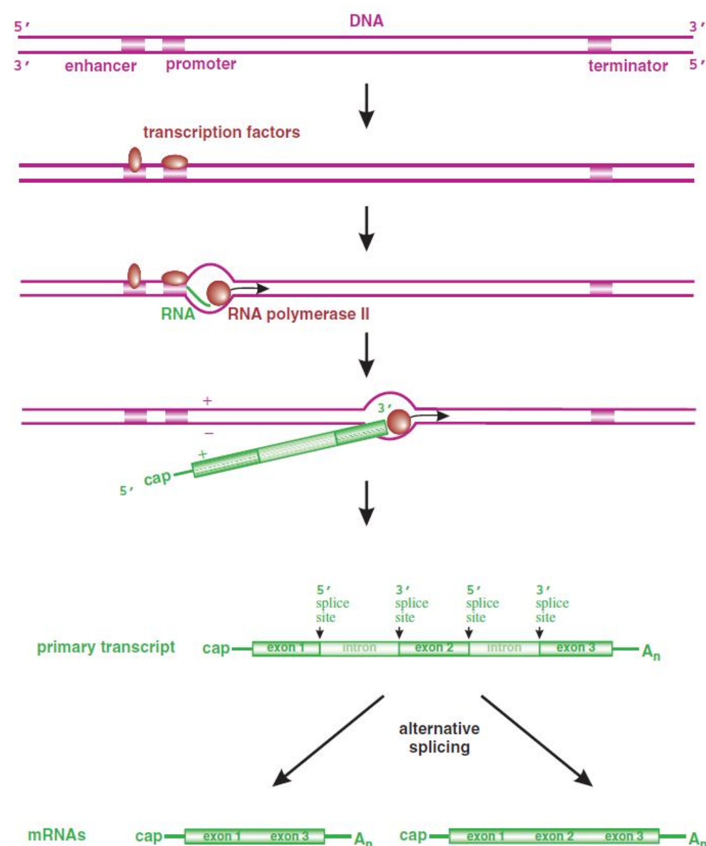

Figure 6.3 Transcription from dsDNA in eukaryotes. Transcription is initiated after transcription factors bind to sequences within promoters and enhancers. The primary transcript is capped at the 5′ end and polyadenylated at the 3′ end. The mRNAs are formed by removal of introns from the primary transcript.

6.3.1 Promoters and enhancers

The following consensus sequence is present in the promoters of many eukaryotic cell and virus genes:
T A T A A/T A A/T A/G

The sequence is known as a TATA box and is usually located 25–30 bp upstream from the transcription start site. A TATA box is present, for example, in the single promoter of HIV-1 (Chapter 17), but in only one of the four promoters of hepatitis B virus (Section 18.8.3).

Enhancers contain sequences that bind transcriptionfactors and these interactions may increase the rate of transcription starts by RNA polymerase II. Remarkably,some cell enhancers are up to 1 Mbp upstream or downstream from their promoters, though an enhancer and a promoter may come into close proximity as a result of DNA folding. Many enhancers are cell-type specific.

6.3.2 Transcription factors

Transcription factors are proteins that bind specifically to promoter and enhancer sequences to control gene expression. Some viruses produce their own transcription factors, such as herpes simplex virus VP16, which is a component of the virion (Section 11.5.2), and human T cell leukaemia virus I Tax protein, which is produced in the infected cell (Section 22.10.2).

Some cell transcription factors can activate or repress transcription of viral genes. Tissue-specific transcription factors are required by some viruses, which probably explains why some viruses are tissue specific. Some cell transcription factors, known as general transcription factors, are involved in controlling the expression of genes in many cell types. An example is transcription factor IID (TFIID), which binds to the TATA box (Figure 6.4). TFIID is a complex of 13 polypeptides, one of which is the TATA box binding protein. After TFIID has bound to the TATA box other general transcription factors (TFIIA, IIB, IIE, IIF and IIH) and RNA polymerase II bind.

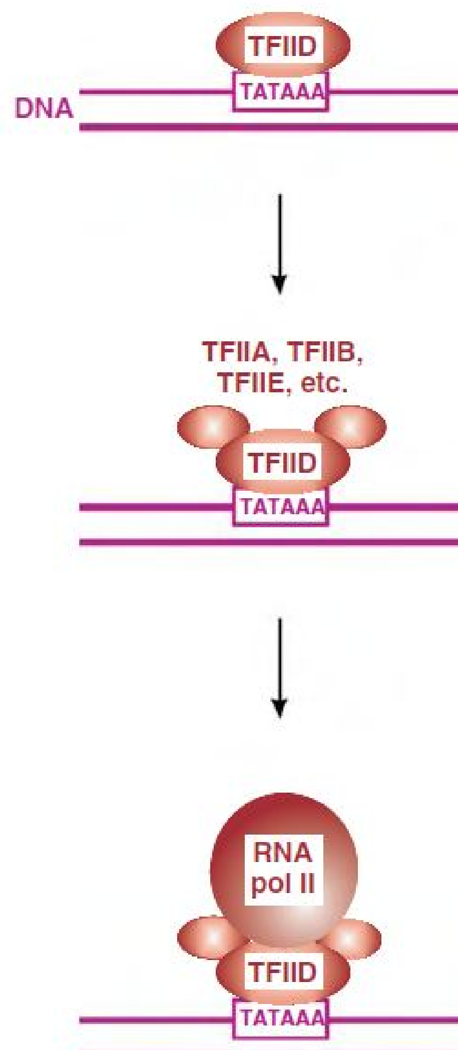

Figure 6.4 Binding of transcription factors and RNA polymerase II at a TATA box.

Among the cell transcription factors that bind to enhancers are

- AP-1 and AP-2 (activator proteins 1 and 2)
- Sp1 (stimulatory protein 1)
- NF-κB (nuclear factor κB).

Most of these transcription factors are involved in HIV-1 transcription (Section 17.4.3).

As well as activating gene expression, transcription factors are also involved in the repression of gene expression. All organisms regulate expression of their genes. A frog has different genes switched on depending on whether it is in the embryo, tadpole or adult stage. Similarly, a virus may have different genes switched on early and late in the replication cycle. For some viruses three phases (e.g. herpesviruses; Section 11.5.2) or four phases (e.g. baculoviruses) of gene expression can be distinguished.

6.3.3 Transcriptases

Transcriptase is a general term for an enzyme that carries out transcription. Viruses that replicate in the nucleus generally use a cell enzyme, while viruses that replicate in the cytoplasm encode their own (Figure 6.5).

A DNA virus needs a DNA-dependent RNA polymerase to transcribe its genes into mRNA. Viruses that carry out transcription in the nucleus generally use the cell RNA polymerase II; these include the retroviruses, as well as many DNA viruses. DNA viruses that replicate in the cytoplasm use a virus-encoded enzyme because there is no appropriate cell enzyme in the cytoplasm.
An RNA virus (apart from the retroviruses) needs an RNA-dependent RNA polymerase to transcribe its genes into mRNA. Each virus in Classes III, IV and V encodes its own enzyme, in spite of the fact that the cells of plants and some other eukaryotes encode ssRNA-dependent RNA polymerases.

All the viruses that carry out transcription in the cytoplasm, except the plus-strand RNA viruses, have the transcriptase in the virion so that the enzyme is immediately available to transcribe the virus genome when a cell is infected. Before the plusstrand RNA viruses can begin transcription they must translate copies of the enzyme from the genome RNA.

The retroviruses and the pararetroviruses perform reverse transcription (Section 6.2) using enzymes known as reverse transcriptases. These enzymes are RNA-dependent DNA polymerases, but they also have DNA-dependent DNA polymerase activity, as the process of reverse transcription involves synthesis of DNA using both RNA and DNA as the template.

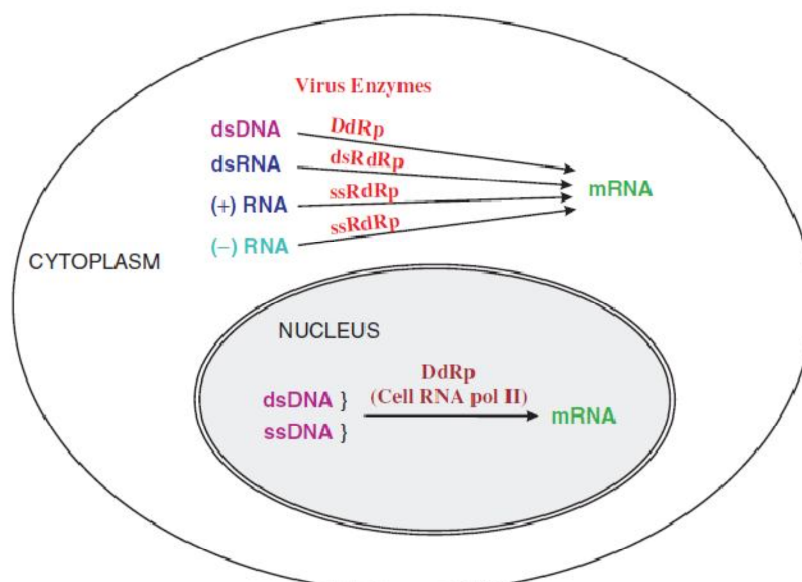

Figure 6.5 Enzymes used by viruses to transcribe their genomes to mRNA. A (+) RNA genome has the same sequence as the virus mRNA. A (–) RNA genome has the sequence complementary to that of the virus mRNA. Transcription from dsDNA in the nucleus applies not only to those dsDNA viruses that replicate in the nucleus, but also to the reverse transcribing viruses.
DdRp: DNA-dependent RNA polymerase; **dsRdRp**: double-stranded RNA-dependent RNA polymerase; **ssRdRp**: single-stranded RNA-dependent RNA polymerase

6.3.4 Capping transcripts

Soon after RNA synthesis has begun, and while transcription is continuing, most transcripts are 'capped' at the 5′ end (Figure 6.3). The cap is a guanosine triphosphate joined to the end nucleotide by a 5′–5′ linkage, rather than the normal 5′–3′ linkage. A methyl group is added to the guanosine, and in some cases to one or both of the ribose residues on the first and second nucleotides (Figure 6.6). Throughout the book we shall use a cartoon cap to depict a cap on the 5′ end of an RNA molecule.

Most eukaryotic cell and viral mRNAs have a cap at their 5′ end. The cap is thought to
- aid mRNA transport from the nucleus to the cytoplasm;
- protect the mRNA from degradation by exonucleases;
- be required for the initiation of translation.

The cell enzymes that carry out the capping activities are guanylyl transferases (they add the guanosine 5′-triphosphate) and methyl transferases (they add the methyl groups). These enzymes are located in the nucleus and most of the viruses that carry out transcription in the nucleus, like the retroviruses, use the cell enzymes. Many of the viruses that replicate in the cytoplasm, however, encode their own capping and methylating enzymes; these viruses include the poxviruses, the reoviruses and the coronaviruses.

Minus-strand RNA viruses with segmented genomes have evolved a mechanism to 'snatch' caps from cell mRNAs. These viruses include animal viruses, such as influenza viruses, and plant viruses, such as tomato spotted wilt virus. The complex of virus proteins making up the RNA polymerase binds to cellular capped mRNA, then an endonuclease activity associated with the complex cleaves the RNA, generally 10–20 nucleotides from the 5′ end. The capped oligonucleotides act as primers to initiate transcription of viral mRNA.

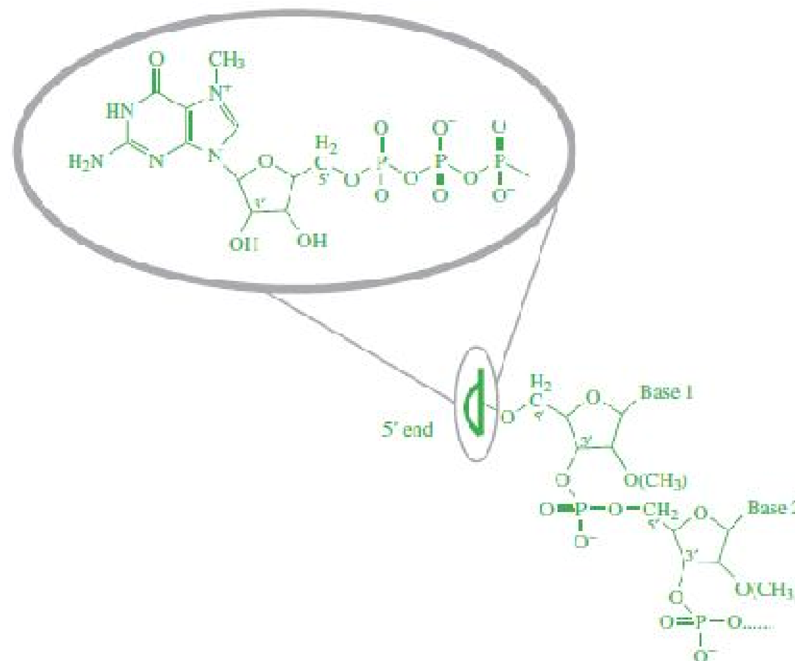

Figure 6.6 Cap on 5′ end of messenger RNA. The inset shows the structure of the cartoon cap, which depicts a methylated guanosine triphosphate linked to the first nucleotide by a 5′–5′ linkage.

Not all mRNAs are capped. Picornaviruses, for example, do not cap their mRNAs; these viruses replicate in the cytoplasm, so their RNAs do not require transport from the nucleus, and translation is initiated by a mechanism that is not dependent upon a cap (Chapter 14).

6.3.5 Polyadenylation of transcripts

A series of adenosine residues (a polyadenylate tail; poly(A) tail) is added to the 3′ end of most primary transcripts of eukaryotes and their viruses. Polyadenylation probably increases the stability of mRNAs, and the poly(A) tail plays a role in the initiation of translation (Section 6.4.1). These functions can be provided in other ways, however, as some viruses, such as the reoviruses (Section 13.3.2), do not polyadenylate their mRNAs.

In most cases there is a polyadenylation signal about 10–30 bases upstream of the polyadenylation site. The polyadenylation signal AATAAA was first characterized in simian virus 40 in 1981. It has since been found that this sequence is used by many other animal viruses, such as HIV-1 (Section 17.4.3) and Rous sarcoma virus, as well as by animal cells. Some viruses use other sequences as polyadenylation signals; for example, the mammalian hepadnaviruses (Section 18.8.3) use TATAAA, a sequence that can function as a TATA box in other contexts!

In most cases the poly(A) tail is added by the following mechanism. During transcription the RNA polymerase proceeds along the template past the polyadenylation signal and the polyadenylation site. The newly synthesized RNA is then cleaved at the polyadenylation site and the poly(A) tail is added step by step by a complex of proteins, including a poly(A) polymerase. Some viruses have evolved alternative mechanisms to polyadenylate their mRNAs; these viruses include the picornaviruses (Section 14.4.4) and the rhabdoviruses (Section 15.4.2).

6.3.6 Splicing transcripts

Some primary transcripts are functional mRNAs, but most eukaryotic cell primary transcripts contain sequences (introns) that are removed. The remaining sequences (exons) are spliced at specific donor sites and acceptor sites to produce the mRNAs (Figure 6.3). A primary transcript may be cut and spliced in more than one way to produce two or more mRNA species. Some primary transcripts of viruses that replicate in the nucleus are processed in the same way to produce the virus mRNAs. The first evidence of split genes, as they are known, was reported in 1977 after studies with adenoviruses.

Further examples of viruses that have split genes are herpesviruses (Section 11.5.2), parvoviruses (Section 12.4.3) and retroviruses (Section 16.3.4). The simplest type of split gene consists of two exons separated by one intron, but some are much more complex; gene K15 of Kaposi's-sarcoma-associated herpesvirus has eight exons and seven introns. The HIV-1 genome has a number of splice donor sites and acceptor sites; splicing of the primary transcript results in more than 30 different mRNA species (Section 17.4.3).

6.4 Translation in eukaryotes

A typical eukaryotic mRNA is monocistronic, i.e. it has one ORF from which one protein is translated (Figure 6.7). Sequences upstream and downstream of the ORF are not translated. Some large ORFs encode polyproteins, large proteins that are cleaved to formtwo or more functional proteins.

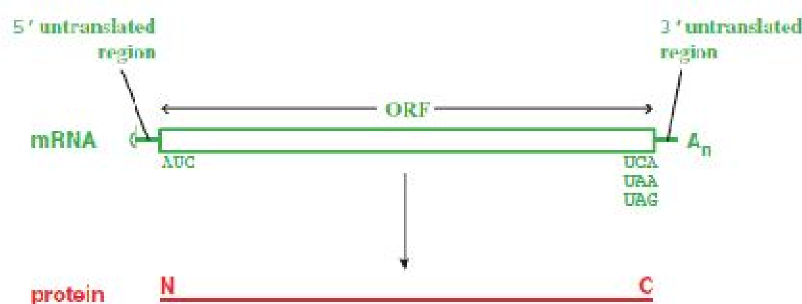

Figure 6.7 Translation from a monocistronic mRNA. There is one open reading frame (ORF), usually starting at the first AUG codon from the 5′ end of the mRNA, and ending at a stop codon (UGA, UAA or UAG). Translation occurs in the 5′ to 3′ direction, the N terminus of the protein being synthesized first.

6.4.1 Initiation of translation

As we have noted, most eukaryotic cell and virus mRNAs have a methylated nucleotide cap at the 5′ end and a poly(A) sequence at the 3′ end. These structures play key roles in the initiation of translation. The cap is especially important; it is the binding site for eukaryotic initiation factors (eIFs), a methionine tRNA charged with its amino acid, and a 40S ribosomal subunit (Figure 6.8). A poly(A)-binding protein binds to the poly(A) tail. The proteins bound at the ends of the RNA are able to interact, and it is thought that this interaction might allow circularization of the mRNA leading to stimulation of translation. Messenger RNAs that lack a cap and/or a poly(A) sequence might be circularized by other mechanisms.

The 40S ribosomal subunit is moved along the RNA in the 5′ to 3′ direction, scanning until an initiation codon is encountered in an appropriate sequence context. The initiation codon is normally AUG, and is normally the first AUG from the 5′ end. Some viruses, however, use other initiation codons; Sendai virus uses ACG as the initiation codon for one of its genes.

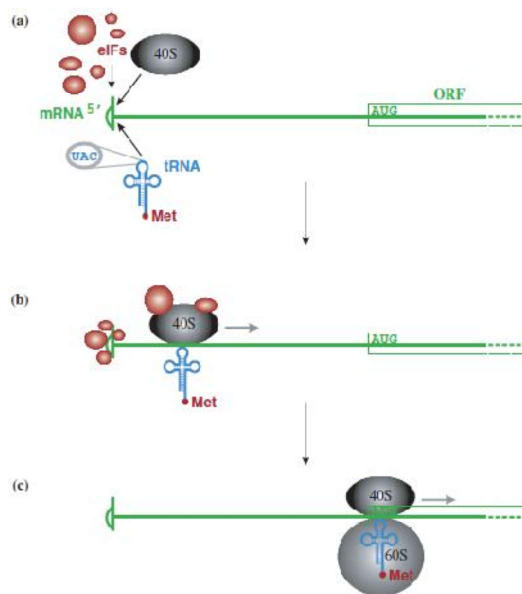

Figure 6.8 Initiation of translation on a capped mRNA. (a) Eukaryotic initiation factors (eIFs), a 40S ribosomal subunit and a methionine linked to its tRNA bind at the 5′ end of the mRNA. (b) The complex scans from the 5′ end of the mRNA. (c) When the first AUG codon is reached it is recognized by the anticodon UAC in the tRNA; a 60S ribosomal subunit is bound and initiation factors are released.

Some mRNAs are not capped and initiation of translation occurs by a different mechanism. A reduced set of eIFs binds not at the at the 5′ end, but at an internal ribosome entry site (IRES), which has a high degree of secondary structure (Figure 6.9).

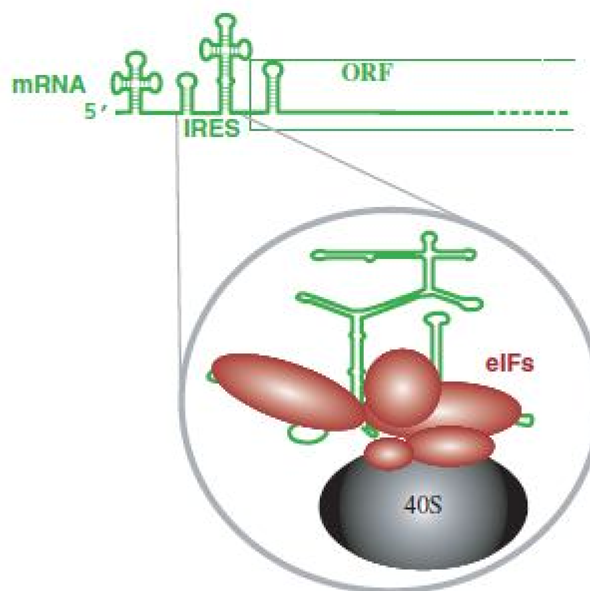

Figure 6.9 Initiation of translation on an mRNA that is not capped. A 40S ribosomal subunit and eIFs bind at an internal ribosome entry site (IRES) upstream of the ORF.

IRESs are present in a number of RNA viruses, including hepatitis C virus and the picornaviruses (Chapter 14). They have also been found in cell mRNAs and in one of the mRNAs of a DNA virus (Kaposi's-sarcoma-associated herpesvirus).

6.4.2 Translation from bicistronic mRNAs

Most eukaryotic cell and virus mRNAs have one ORF, but there are a number of virus mRNAs that have two or more ORFs. Some of these bicistronic and polycistronic mRNAs are functionally monocistronic, but some structurally bicistronic mRNAs are functionally bicistronic. A difference in the rate of translation of the two ORFs provides a mechanism for expressing two genes at different levels.

In many bicistronic mRNAs the ORFs overlap (Figure 6.10(a)); in others there is an ORF within an ORF (Figure 6.10(b)). One mechanism to read the second ORF involves leaky scanning; a 40S ribosomal subunit may overlook the ORF 1 start codon and initiate translation at the start of ORF 2. The ORFs for the two proteins are in different reading frames, so the proteins that they encode are unrelated. Of course it is essential that the sequence 'makes sense' in both reading frames!

Another mechanism for reading a second ORF in an mRNA involves ribosomal frameshifting; a ribosome shifts into a different reading frame towards the end of ORF 1. It therefore does not recognize the ORF 1 stopcodon, but continues along the mRNA, reading ORF 2 to produce an elongated version of the ORF 1 protein(Figure 6.10(c)). Frameshifting occurs when the ribosome moving along the RNA encounters a frameshift signal (a specific sequence) followed by a secondary structure, usually a pseudoknot (Section 3.2.2).

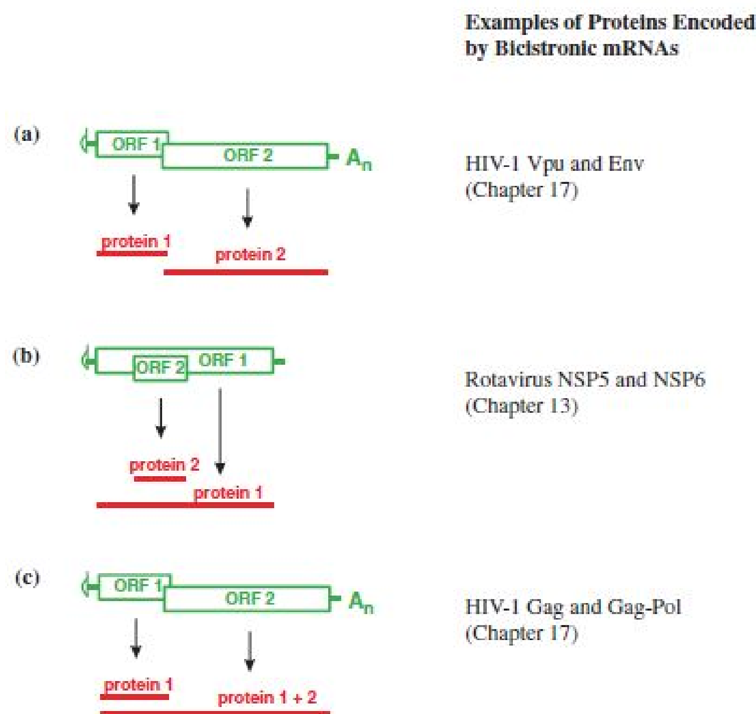

Figure 6.10 Translation from bicistronic mRNAs. (a), (b) A ribosome may begin translation at the start of ORF 1, or scanning may be 'leaky' and translation may begin at the start of ORF 2. The two start codons are in different reading frames and the two proteins are unrelated. (c) ORF 2 is translated by a ribosomal frameshift, producing an extended version of protein 1.

6.4.3 Co- and post-translational modification of proteins

During and after their translation proteins may undergo one or more modifications, including glycosylation, acylation and phosphorylation.

6.4.3.a Glycosylation

Glycosylation involves the addition of oligosaccharide groups to the polypeptide chain. When an oligosaccharide is linked through the –OH group of a serine or threonine residue the process is known as O-glycosylation; when it is linked through the –NH2 group of an asparagine residue it is known as N-glycosylation.

Proteins destined for glycosylation are synthesized in the rough endoplasmic reticulum, where Nglycosylation commences. They are then transported to the Golgi complex (Section 6.5), where N-glycosylation is completed by enzymes such as α-mannosidases I and II and galactosyl transferase. O-glycosylation takes place in the Golgi complex.

Some glycoproteins have undergone only one type of glycosylation, such as the N-glycosylated gp120 of HIV-1, while many glycoproteins have undergone both O- and N-glycosylation, such as gC and gD of herpes simplex virus.

These glycoproteins of HIV and herpes simplex virus, like most virus glycoproteins, are integral membrane proteins that are components of the virion envelopes. Some glycoproteins, however, are not associated with virion envelopes; rotaviruses have naked virions but their surface protein (VP7) is glycosylated (Chapter 13). A few virus glycoproteins are nonstructural proteins, such as the rotavirus protein NSP4.

6.4.3.b Acylation

Acylation is the addition of an acyl group (R–CO–) to a molecule. An acyl group that is commonly added to proteins is a myristyl group, where R is CH3–(CH2)12. The myristyl group is linked to a glycine residue at the N terminus of the protein. Most viruses lack the enzyme N-myristyltransferase that is required for this modification; if one or more of their proteins is myristylated the process is carried out by a host enzyme. Many myristylated proteins associate with membranes. This is true for the Gag proteins of most retroviruses (Chapter 16); if these proteins are not myristylated they do not associate with the plasma membrane and virion assembly does not take place. Another example of a myristylated protein is the picornavirus capsid protein VP4 (Chapter 14).

6.4.3.c Phosphorylation

Phosphorylation involves the transfer of a phosphate group from a nucleotide, usually ATP, to the O of an –OH group of a serine, threonine or tyrosine residue. The transfer is carried out by protein kinases, which may be of cell and/or viral origin. The enzymes recognize short sequences of amino acids that bracket the residue to be phosphorylated.

Phosphorylation can alter the conformation, activity,localization and/or the stability of a protein, and many cell and viral processes involve protein phosphorylation. Many structural and non-structural virus proteins become phosphorylated, for example the phosphoproteins of rhabdoviruses (Chapter 15); one-sixth of the amino acid residues in the phosphoprotein of vesicular stomatitis virus are serine and threonine, and many of these are phosphorylated.

6.5 Transport in eukaryotic cells

We have already discussed how nucleocapsids and other virus structures are transported via microtubules and nuclear pores after entry into a cell (Section 5.2.5). Virus molecules synthesized in the infected cell must also be transported to particular sites. Virus mRNAs are transported from the nucleus to the cytoplasm, and virus proteins may be transported to various locations, including the nucleus (Figure 6.11).

Many proteins have a sequence of amino acids (a 'post code') that specifies their destination. Proteins destined to be incorporated into membranes have a signal sequence, which is a series of hydrophobic amino acid residues, either at the N terminus or internally. Protein synthesis begins on a free ribosome, but when the signal sequence has been synthesized it directs the polypeptide–ribosome complex to the endoplasmic reticulum, where protein synthesis continues. Regions of the endoplasmic reticulum with ribosomes associated are known as rough endoplasmic reticulum (Figure 6.11).

Each integral membrane protein has one or more membrane anchor sequences, which are rich in hydrophobic amino acid residues. For some of these proteins the signal sequence acts as a membrane anchor. Other integral membrane proteins, like the HIV-1 envelope protein, are moved through the membrane until the anchor sequence is reached, and then the signal sequence is removed by a host enzyme.

Many of the proteins synthesized in the rough endoplasmic reticulum are transported via vesicles to the Golgi complex, and most integral membrane proteins become glycosylated in these membrane compartments. From here the glycoproteins may be transported to other membranes, such as the plasma membrane or the nuclear envelope. Progeny virions may bud from these membranes (Section 8.3.1).

Epithelial cells have apical (outer) and basolateral (inner) surfaces, which are composed of different lipids and proteins and are separated by 'tight junctions'. During infections of epithelial cells with enveloped viruses budding of virions from the plasma membrane may be restricted to either the apical surface or the basolateral surface.

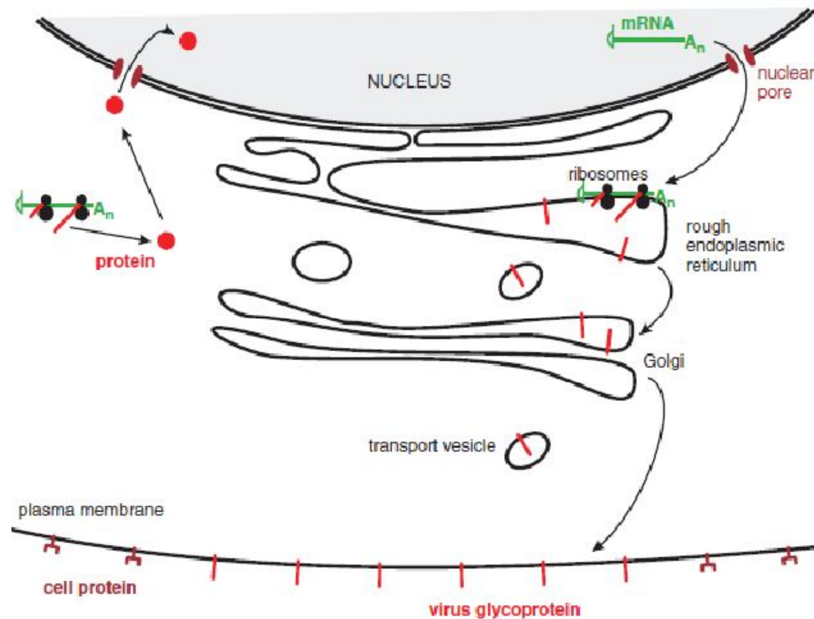

Figure 6.11 Synthesis and intracellular transport of molecules. Some proteins translated on free ribosomes are transported via nuclear pores to the nucleus. Some proteins translated in the rough endoplasmic reticulum are transported via the Golgi complex to the plasma membrane.

Enrique Rodriguez-Boulan and David Sabatini used electron microscopy to examine thin sections of virus infected epithelial cell cultures that retained their differentiated surfaces. They reported in 1978 that vesicular stomatitis virus buds from the basolateral surface, while influenza virus buds from the apical surface. This probably explains why most influenza virus infections of mammals are localized to the respiratory tract. It can be demonstrated that the glycoproteins of each virus are targeted to the appropriate surface (Figure 6.12). The targeting signal sequences and their locations within the proteins have been determined.

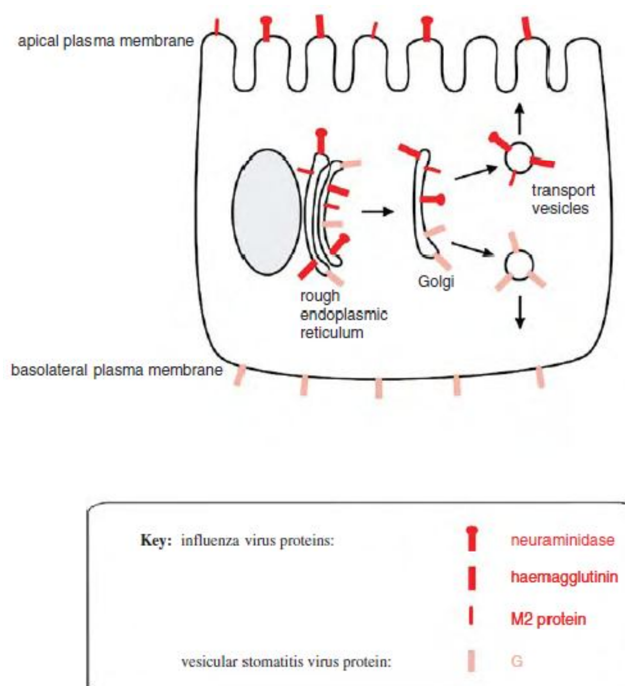

Figure 6.12 Targeting of virus envelope proteins to apical and basolateral surfaces of an epithelial cell. Influenza A virus and vesicular stomatitis virus envelope proteins are transported to the apical and basolateral surfaces, respectively.

If the virus replicates in the nucleus then most, if not all, of the virus proteins must be transported into the nucleus. These proteins, like cell proteins that are transported into the nucleus, have a nuclear localization signal, which is rich in one or both of the basic amino acids lysine (K) and arginine (R). A common nuclear localization signal is

PKKKRKV.

This signal was first identified in a simian virus 40 protein known as large T antigen. A nuclear localization signal allows a protein to bind to cell proteins (importins) and subsequently to nuclear pore filaments (Figure 5.6), from where it is transported into the nucleus. RNAs are also transported within the cell; for example, mRNAs synthesized in the nucleus must be exported through nuclear pores to the cytoplasm. The RNAs are taken to their destinations by proteins. The Rev protein of HIV-1 has both a nuclear localization signal and a nuclear export signal (Figure 6.13). The nuclear localization signal ensures that Rev is transported into the nucleus, where it binds specifically to HIV-1 RNA. The nuclear export signal ensures that Rev and its RNA cargo are transported from the nucleus to the cytoplasm via a nuclear pore.

Figure 6.13 HIV-1 Rev protein. The nuclear localization signal is rich in arginine (R) residues. The nuclear export signal is rich in leucine (L) residues.

6.6 Transcription and translation in bacteria

This section starts with a brief summary of transcription from prokaryotic genes. There is one type of DNA-dependent RNA polymerase for transcription in prokaryotes, and the E. coli holoenzyme comprises the catalytic core enzyme containing one β, one β′ and two α subunits, and a regulatory protein, the sigma (σ) factor. The σ factor determines promoter specificity, enabling the polymerase to recognize and bind to specific promoters in the correct orientation to initiate transcription. Bacterial promoters contain characteristic sequences: the −10 region (Pribnow box) and the −35 region, centred 10 and 35 bp upstream of the transcription start site respectively, and recognized by the σ factor. Upstream of the −35 region some promoters additionally contain the UP element, which is recognized by the RNA polymerase α subunit and enhances binding of the enzyme.

Once transcription initiation is completed, the σ factor dissociates to be re-used in the process. Different σ factors recognize different promoters. There are two main sigma families: σ70 and σ54. The σ70 family has many different σ factors, including the primary (vegetative) σ factors, directing transcription of genes for bacterial growth and metabolism, and alternative factors, e.g. σ32 for transcription of heat shock genes.

Transcription terminates after a terminator is transcribed. Termination may be rho independent (intrinsic) or rho dependent. Rho-independent termination depends on the template DNA sequence, which contains an inverted repeat and a string of adenine residues, located downstream of the rho-independent genes. The inverted repeat forms a stem-loop structure in the RNA transcript and adenine–uracil bonds form in the DNA–RNA hybrid, such that the RNA transcript can dissociate from the template. Rho-dependent termination involves a rho factor, which has helicase and ATPase activity, to unwind DNA–RNA hybrids when the polymerase is stalled at an inverted repeat. This leads to release of the transcript.

Bacterial cells and the viruses that infect them operate transcription and translation mechanisms that differ in a number of respects from those in eukaryotes. Introns are rarely present in the genes of bacteria and their viruses. Some phages use the host DNAdependent RNA polymerase for their transcription, while others encode their own.

Whereas typical mRNAs in eukaryotes are monocistronic, in bacteria they are polycistronic, i.e. each mRNA has several ORFs (Figure 6.14). The mRNAs of bacteria and their viruses are never capped at the 5′ end, and are rarely polyadenylated at the 3′ end.

Figure 6.14 Characteristics of bacterial and phage mRNA. There are typically several ORFs, all of which may be translated at the same time. The 5′ end is not capped and the 3′ end is rarely polyadenylated.

Bacterial translation differs from eukaryotic translation in a number of features.

- Translation may start before transcription is complete. The lack of a nucleus allows transcription and translation to be coupled.
- The ribosomes are smaller (the ribosomal subunits have sedimentation coefficients of 30S and 50S).
- The 30S ribosomal subunit binds directly to a translation initiation region on the mRNA. Initiation generally involves interaction of the Shine- Dalgarno (S-D) sequence at the ribosome binding site (RBS) on mRNA and the anti-S-D sequence at the 3′ end of 16S rRNA in the 30S subunit. The RBS is located just upstream of the AUG start codon.
- The methionine of the initiator methionyl tRNA is generally formylated.
- A much smaller number of initiation factors is involved.
- All ORFs within an mRNA are translated and several may be translated concurrently.

A few phages have overlapping genes, which may be translated by reading through a stop codon or by ribosomal frameshifting (Section 6.4.2).

===================================hendropramono====================================

Orang besar bukan orang yang otaknya sempurna tetapi orang yang mengambil sebaik-baiknya dari otak yang tidak sempurna

Berusahalah untuk tidak menjadi manusia yang berhasil tapi berusahalah menjadi manusia yang berguna

Bukan kecerdasan anda, melainkan sikap andalah yang yang akan mengangkat anda dalam kehidupan

Orang yang berjiwa besar teguh pendiriannya, tetapi tidak keras kepala